Predicting the Probability of Long-Term Unemployment in Ireland Using Administrative Data

Seamus McGuinness
Elish Kelly
John R. Walsh

ESRI SURVEY AND STATISTICAL REPORT SERIES
NUMBER 51

June 2014

ESRI

Predicting the Probability of Long-Term Unemployment in Ireland Using Administrative Data

Seamus McGuinness
Elish Kelly
John R. Walsh

ESRI SURVEY AND STATISTICAL REPORT SERIES

NUMBER 51

June 2014

Available to download from *www.esri.ie*

© The Economic and Social Research Institute
Whitaker Square, Sir John Rogerson's Quay, Dublin 2

ISBN 978 0 7070 03696

The ESRI

The *Economic Research Institute* was founded in Dublin in 1960, with the assistance of a grant from the Ford Foundation of New York. In 1966 the remit of the Institute was expanded to include social research, resulting in the Institute being renamed *The Economic and Socia*l *Research Institute* (ESRI). In 2010 the Institute entered into a strategic research alliance with Trinity College Dublin, while retaining its status as an independent research institute.

The ESRI is governed by an independent Council which acts as the board of the Institute with responsibility for guaranteeing its independence and integrity. The Institute's research strategy is determined by the Council in association with the Director and staff. The research agenda seeks to contribute to three overarching and interconnected goals, namely, economic growth, social progress and environmental sustainability. The Institute's research is disseminated through international and national peer reviewed journals and books, in reports and books published directly by the Institute itself and in the Institute's working paper series. Researchers are responsible for the accuracy of their research. All ESRI books and reports are peer reviewed and these publications and the ESRI's working papers can be downloaded from the ESRI website at www.esri.ie

The Institute's research is funded from a variety of sources including: an annual grant-in-aid from the Irish Government; competitive research grants (both Irish and international); support for agreed programmes from government departments/agencies and commissioned research projects from public sector bodies. Sponsorship of the Institute's activities by Irish business and membership subscriptions provide a minor source of additional income.

The Author

Seamus McGuinness is an Associate Research Professor, Elish Kelly is a Research Officer and John R. Walsh is a Senior Research Analyst at the Economic and Social Research Institute. Dr McGuinness and Dr. Kelly hold adjunct research positions at Trinity College Dublin, and Dr. Seamus McGuinness is also adjunct Associate Professor at the National Institute of Labor Studies, Flinders University Australia.

The paper has been accepted for publication by the Institute, which does not itself take institutional policy positions. ESRI Survey and Statistical Reports are peer reviewed by ESRI research colleagues. The authors are solely responsible for the content and the views expressed.

Acknowledgements

The work carried out in this report was funded by the Department of Social Protection and we would like to thank all individuals within the Department who provided assistance during the project, along with valuable comments on earlier drafts of the report. In particular, we would like to thank Barry Kennedy, Paul Morrin and Brian King for their active engagement with the project and assistance with queries and access to data. Thanks also to the internal reviewer who provided comments on an earlier draft, and to Regina Moore of the ESRI for all her work in preparing the manuscript for publication.

Table of Contents

List of Tables

List of Figures

1. Background and Introduction

This report develops a statistical profiling model to be applied to the existing population of long-term unemployed claimants in Ireland who have not yet been profiled through the Department of Social Protection's new Probability of Exit (PEX) profiling model. The PEX model was rolled out as part of the Government's new integrated employment and support service Intreo, which was introduced in October 2012.[1] As with the development of the previous profiling model, this study will provide a basis for scoring individuals based on a set of observable characteristics. However, the current study differs from that undertaken for the development of the PEX model (see O'Connell *et al.*, 2009) in a number of important ways. First, the current study is based on administrative data only, whereas the PEX model was developed using a combination of administrative and profiling survey data. As a consequence, the range of explanatory variables included within the model developed in this study is more restricted. Second, the individuals to which the model developed within this study is being applied to have already reached 12 months duration on the Live Register. This means that the score can no longer be interpreted as a predicted probability of becoming long-term unemployed. Instead, the score can be viewed as the predicted probability that the individual should have already left the Live Register given their characteristics. Thus, the score is more a measure of relative Labour Market Disadvantage (LMD) as opposed to an expected probability of future exit from the Live Register.

[1] The roll out of the PEX model in local Department of Social Protection (DSP) offices had begun before this and the process was completed during 2012.

2. Data and Methods

The Department of Social Protection (DSP) provided the Economic and Social Research Institute (ESRI) with the data required to develop a profiling model of relative Labour Market Disadvantage (LMD). The data came from the Department's new *Jobseekers Longitudinal Dataset*, which tracks individuals' claims histories over time. For the purposes of the study, the DSP provided anonymised claims histories' of all individuals who registered as unemployed in Ireland between 7 March and 16 May 2012. The claim history data then enabled the ESRI to estimate the probability that claimants would exit to the labour market prior to becoming long-term unemployed between 7 March and 16 May 2013. The development of the DSP's *Jobseekers Longitudinal Dataset* substantially simplified the model development process relative to the development of the earlier PEX model. In contrast to this earlier study, the provision of longitudinal information on individual claimants meant that it was not necessary to scan Live Register records for each week of the study period in order to observe individual exits to the labour market.

Anonymised data was received for 63,795 individuals who made claims for Jobseekers Benefit (JB) or Jobseekers Allowance (JA)[2] during the relevant period and who subsequently had their claims approved. With respect to the information held for each individual, this varied somewhat. Basic Live Register information on gender, age, previous occupation, marital status, spousal income, nationality and geographical information is held for all claimants. Additional information on a range of educational and labour market attributes is held in the *Jobseekers Longitudinal Dataset* for approximately 63 per cent of the sample. The basic Live Register data is captured in the Department's ISTS system, whereas the richer information on education and labour market characteristics is captured in the Client Services System (CSS), both of which are components of the *Jobseekers Longitudinal Dataset*. However, it would appear that the level of data captured on individuals is non-random and is related to their unemployment history. CSS data tend to be collected on individuals who had contact with the activating authorities i.e., the DSP and FÁS,[3] whereas individuals for whom the more basic ISTS data exist tend to be relatively new claimants with lower levels of historical contact with either the DSP or FÁS. The ISTS data would also include seasonal, casual and part-time workers (i.e., those who work up to three days a week): these are individuals who are not referred for activation and who, therefore, will not have their information captured in the CSS.

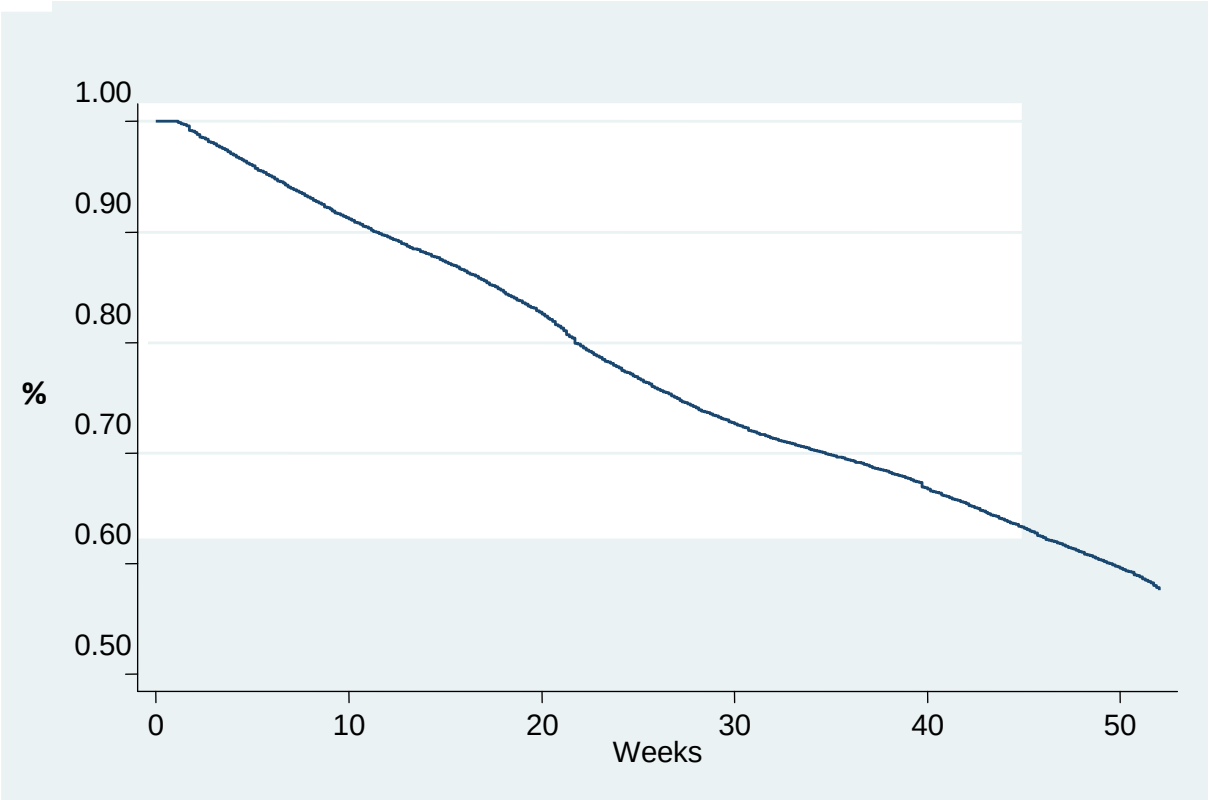
[2] JB is a contributions based benefit, while JA is a means tested benefit.

[3] FÁS was formally Ireland's national employment and training authority but it was disbanded and replaced by a new further education and training body called SOLAS in October 2013. As part of this process, the Community Employment and Employment Services component of FÁS was transferred to the DSP. Former FÁS Training Centres, which were the other main component of FÁS, are currently being transferred to the new Education and Training Boards (ETBs); this process will be completed during 2014. The ETBs are now responsible for the management of training provision.

3. Descriptives

Figure 3.1 plots the Kaplan-Meier (KM) survival function for the entire sample: this function calculates the fraction of individuals exiting the Live Register during successive weeks in unemployment. We can see from this figure that approximately 58 per cent of the total sample was still on the Live Register at the 12 month point (52 weeks), with the exit pattern linear in nature.[4] In Figure 3.2, the sample is restricted to labour market closures[5] only (as opposed to all closures): when we look at the KM survival function for this outcome, we can see that almost 30 per cent exit the Live Register. Comparing the KM survival function for all closures (Figure 3.1) with that of labour market closures only (Figure 3.2), it is obvious that a high proportion of individuals on the Live Register close for non-labour market reasons, e.g., education, training, transfers to other benefits, etc.

Figure 3.1: Kaplan-Meier Survival Estimates - All Closures



[4] This was also the case for the exit functions plotted within the PEX study (O'Connell *et al.*, 2009).

[5] This is a closure to employment.

Figure 3.2: Kaplan-Meier Survival Estimates – Labour Market Closures

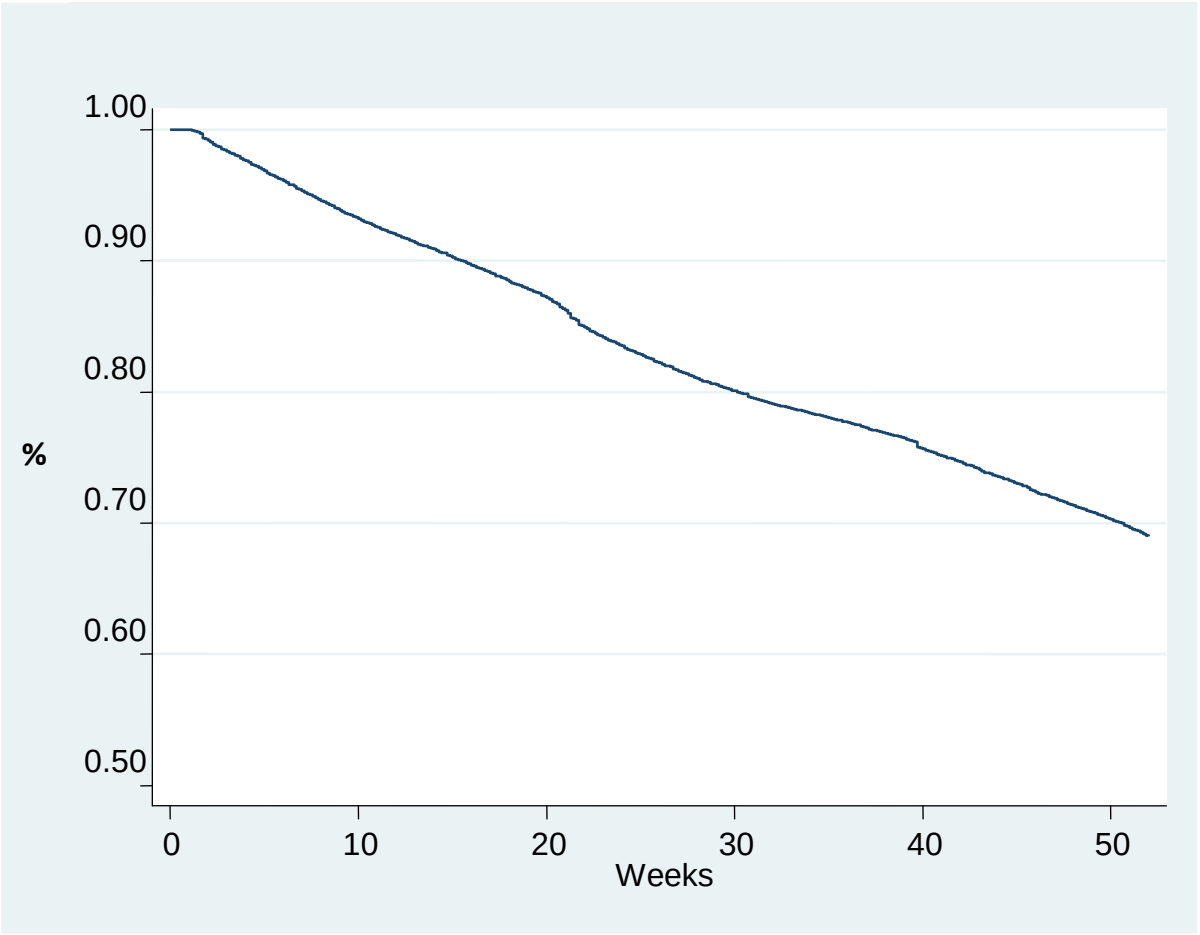


Figure 3.3 shows the KM survival function that is derived for labour market closures in the CSS sample, while the KM survival function presented in Figure 3.4 is based on the ISTS sample. We can see from these two figures that there is a substantial difference between the CSS and the ISTS samples. Based on the CSS data, 73 per cent of the sample had not closed to the labour market by 12 months. This compares with just over 60 per cent of claimant in the ISTS data.

Figure 3.3: Kaplan-Meier Survival Estimates – Labour Market Closures (CSS Sample)

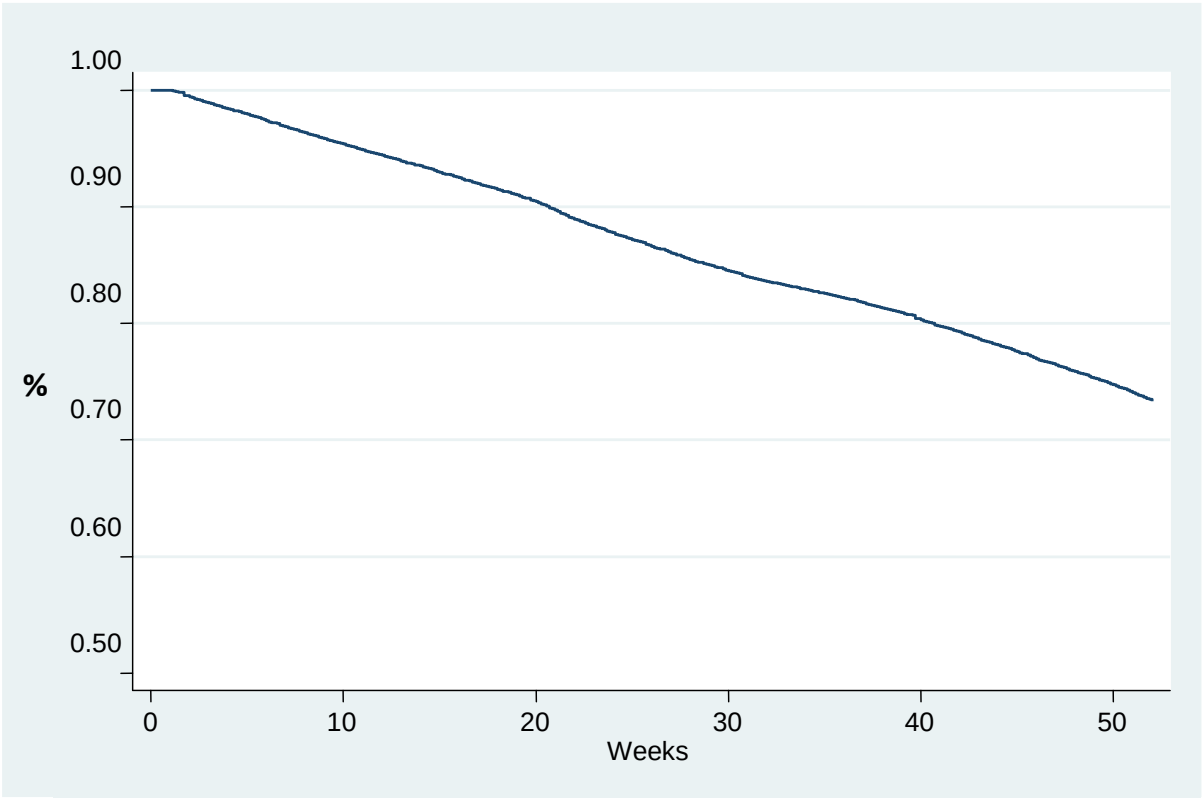


Figure 3.4: Kaplan-Meier Survival Estimates – Labour Market Closures (ISTS Sample)

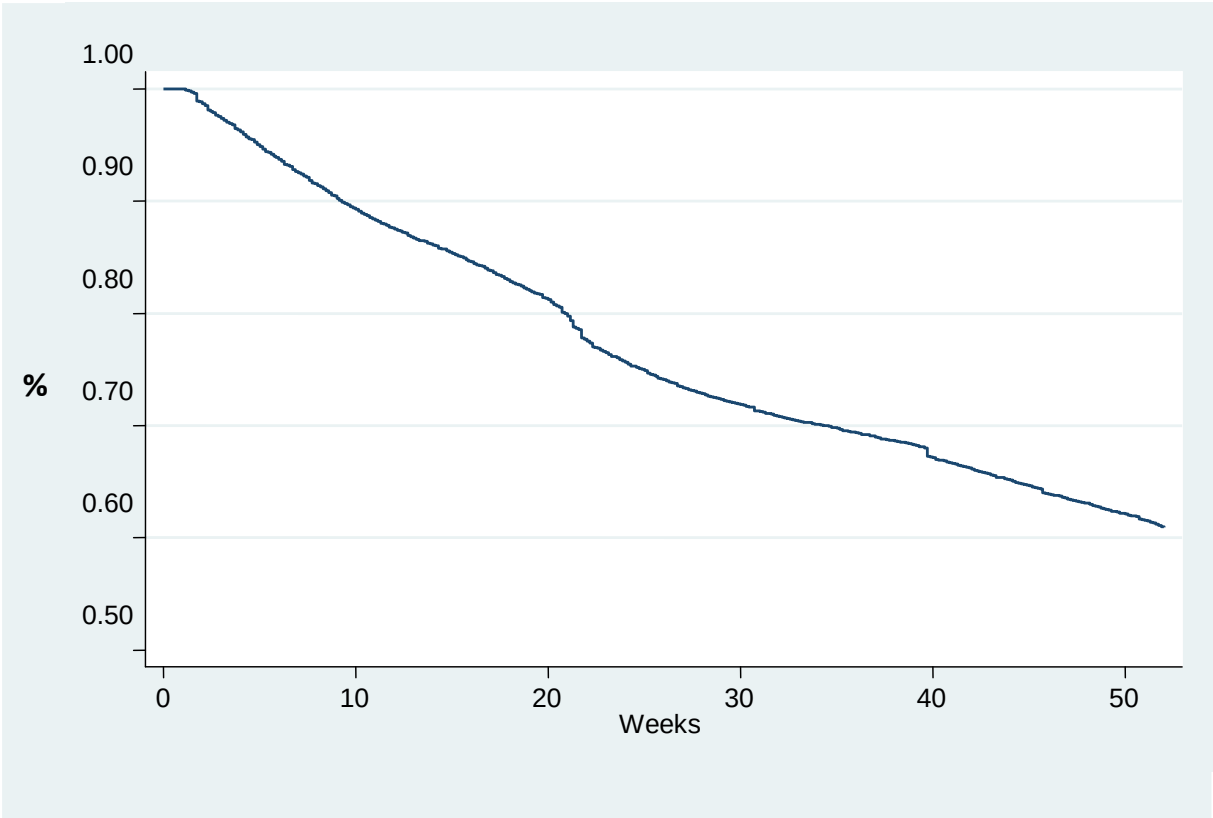


Table 3.1 presents the average incidence of key characteristics, such as age, gender, nationality, spousal earnings, etc., for both the ISTS and CSS samples. In relation to marital status, we can see that a higher proportion of the ISTS claimants were married, whereas a slightly larger proportion of the CSS sample was single. Regarding gender, 64 per cent of the CSS sample was male, which compares with 57 per cent of the ISTS sample. A striking difference between both the ISTS and CSS samples is in relation to benefit type: 60 per cent of ISTS claimants are in receipt of JB compared to only 36 per cent of CSS individuals. This confirms the view that the ISTS sample differs substantially from the CSS sample in terms of recent labour market attachment, which explains the higher rate of labour market exit among ISTS individuals. Both the ISTS and CSS samples have similar nationality distributions (predominately Irish), while a slightly higher proportion of the ISTS sample have a spouse that earns €400 and above per week.

Table 3.1: Summary Means for ISTS and CSS Samples

	ISTS Sample	CSS Sample
Marital Status:		
Married	0.40	0.35
Cohabits	0.06	0.08
Separated	0.04	0.04
Single	0.50	0.53
Gender:		
Male	0.57	0.64
Female	0.43	0.36
Age	36.7	34.9
Nationality:		
Irish	0.79	0.78
Non-Irish	0.21	0.22
Unemployment Benefit:		
Jobseeker's Benefit	0.60	0.36
Jobseeker's Allowance	0.40	0.64
Spousal Earnings:		
No Spousal Earnings	0.83	0.88
€310.00	0.01	0.01
€310.01 - €400.00	0.01	0.01
€400.01 and Above	0.15	0.10

Table 3.1: Summary Means for ISTS and CSS Samples (continued)

	ISTS Sample	CSS Sample
Previous Occupation:		
Manager	0.06	0.05
Professional	0.11	0.07
Associate Professional	0.03	0.03
Clerical	0.13	0.11
Skilled Trades	0.21	0.22
Other Services	0.10	0.10
Sales	0.10	0.11
Operatives	0.17	0.16
Elementary	0.09	0.15
Location:		
Carlow	0.02	0.01
Cavan	0.02	0.01
Clare	0.02	0.02
Cork	0.14	0.09
Donegal	0.04	0.04
Galway	0.05	0.04
Kerry	0.03	0.03
Kildare	0.05	0.04
Kilkenny	0.01	0.01
Laois	0.02	0.02
Leitrim	0.01	0.01
Limerick	0.04	0.05
Longford	0.02	0.01
Louth	0.04	0.04
Mayo	0.03	0.03
Dublin	0.23	0.31
Meath	0.03	0.02
Monaghan	0.01	0.02
Offaly	0.02	0.02

Table 3.1: Summary Means for ISTS and CSS Samples (continued)

	ISTS Sample	CSS Sample
Location:		
Roscommon	0.01	0.01
Sligo	0.01	0.01
Tipperary	0.05	0.04
Waterford	0.03	0.03
Westmeath	0.02	0.02
Wexford	0.03	0.04
Wicklow	0.02	0.03
Sample Size:	14,169	24,353

With respect to previous occupation, a larger proportion of the ISTS sample were previously employed as a Professional, while a larger number of the CSS sample were former Elementary workers. Finally, the geographical distribution of both groups is quite similar, apart from a higher proportion of the ISTS sample dwelling in Cork and a larger number of the CSS group living in Dublin.

4. Labour Market Disadvantage Model

In terms of our methodology, we have taken the view that the CSS marker contains important information on the claimants as individuals included within this dataset are likely to be further from the labour market and tend to have worked previously in non-professional occupations. Rather than estimate separate models for the CSS and ISTS samples, we have taken the approach of estimating a pooled model that includes a variable indicating the presence of CSS data as a separate characteristic. Table 4.1 presents the results for our Labour Market Disadvantage (LMD) models, which have been estimated separately for males and females. Specifically, the result for each characteristic tells us the impact that the particular attribute will have on an unemployed person's likelihood of exiting to employment at 12 months. In terms of our estimation strategy, consistent with our previous profiling study (see O'Connell *et al.*, 2009), we employ probit analysis, exclude from the sample individuals whose reason for closure is unknown and those who closed and re-entered the dataset.[6] The dependant variable is binary and indicates that the claim has closed to employment and has remained closed for at least six weeks.

Focussing on the male results first, we can see that those that cohabit are less likely to have exited to the labour market relative to single males. The likelihood of exiting to employment also decreases with age for males. Interestingly, Irish males are less likely to have exited to the labour market before 12 months relative to non-Irish males. With respect to benefit type, males in receipt of JB were 27 per cent more likely to have exited to the labour market at 12 months relative to those in receipt of JA. The CSS record variable is also highly significant with males categorised in this dataset almost 13 per cent less likely to exit to employment relative to their ISTS-only counterparts. Males whose spouse's earned in excess of €400 per week were 7 per cent more likely to have exited to the labour market at 12 months. Apart from Other Services, males in one of the other 7 previous occupation categories (e.g., Managers, Professional, Associate Professionals, etc.) are more likely to have exited to employment relative to males whose previous occupation was an Elementary job, with the marginal impact highest for claimants previously employed in Professional occupations. Finally some geographical differences exist: relative to Dublin, males from Galway and Kerry, and Westmeath to a lesser extent, were less likely to exit to employment at 12 months; whereas those from Clare, Cork, Kildare, Limerick, Meath, Tipperary and Wexford, and Wicklow to a smaller extent, being more likely to exit to employment.

In relation to the female LMD model, in contrast to the male model, all marital status categories relative to those that are single are less likely to have exited to employment at 12

[6] This exclusion is applied as many of the apparent closures will be as a consequence of administrative factors.

months. Benefit type and being included in the CSS dataset where again found to be important predictors of exit with the marginal effects similar to those in the male model. However, spousal earnings had a differential gender effect with females with high earning spouses (i.e., earning in excess of €301 per week) being less likely to exit to the labour market at 12 months[7]. The impact of geographic location varied, to some extent, by gender as well. As with males, claimants from Kerry and Westmeath were less likely to exit to the labour market at 12 months compared to their counterparts in Dublin. The same was true as well for females from Laois, Mayo, Offaly and Wexford. On the other hand, females from Longford were found to be more likely to exit to a job at 12 months compared to females in Dublin. Finally in relation to previous occupation, similar results that were derived in the male model for the impact of being previously employed as a Manager, Professional, Associate Professional, Clerical or Operative Worker on the likelihood of exiting to employment at 12 months also emerged in the female model. However, the marginal impact for Professional and Managerial occupations was somewhat larger in the female equation. Unemployed females that were previously employed in Other Services were also found to be more likely to find a job at 12 months relative to former Elementary job workers.

Table 4.1: Exits to the Labour Market at 12 Months (Marginal Effects)

	Male	Female
Marital Status (Ref: Single)		
Married	-0.0064	-0.1437***
	(0.009)	(0.012)
Cohabits	-0.0700***	-0.1569***
	(0.014)	(0.016)
Separated	-0.0000	-0.1275***
	(0.020)	(0.021)
Age	-0.0085***	-0.0056***
	(0.000)	(0.000)
Nationality (Ref: Non-Irish)		
Irish	-0.0465***	-0.0595***
	(0.008)	(0.011)
Unemployment Benefit Type (Ref: Jobseeker's Allowance)		
Jobseeker's Benefit	0.2694***	0.2476***
	(0.007)	(0.010)

[7] This result in consistent with our earlier profiling study (see O'Connell *et al.*, 2009).

Table 4.1: Exits to the Labour Market at 12 Months (Marginal Effects) (continued)

	Male	Female
CSS Record	-0.1262***	-0.1469***
	(0.007)	(0.009)
Spousal Earnings (Ref: No Spousal Earnings)		
€310.00	-0.0454	-0.0433
	(0.037)	(0.041)
€310.01 - €400	0.0090	-0.1231**
	(0.038)	(0.058)
€400.01 and Above	0.0709***	-0.0291**
	(0.014)	(0.014)
Previous Occupation (Ref: Elementary)		
Manager	0.0621***	0.1542***
	(0.018)	(0.023)
Professional	0.2173***	0.2539***
	(0.016)	(0.018)
Associate Professional	0.1127***	0.1119***
	(0.022)	(0.026)
Clerical	0.1116***	0.1649***
	(0.018)	(0.018)
Skilled Trades	0.1001***	-0.0005
	(0.011)	(0.026)
Other Services	0.0257	0.0337*
	(0.017)	(0.018)
Sales	0.0642***	0.0028
	(0.015)	(0.018)
Operatives	0.0930***	0.0539***
	(0.012)	(0.020)

Table 4.1: Exits to the Labour Market at 12 Months (Marginal Effects) (continued)

	Male	Female
Location (Ref: Dublin)		
Carlow	0.0302	0.0475
	(0.030)	(0.036)
Cavan	-0.0408	-0.0005
	(0.026)	(0.035)
Clare	0.0535**	0.0023
	(0.026)	(0.030)
Cork	0.0562***	-0.0013
	(0.012)	(0.016)
Donegal	0.0159	0.0350
	(0.018)	(0.024)
Galway	-0.0561***	-0.0316
	(0.017)	(0.022)
Kerry	-0.0419**	-0.0717***
	(0.021)	(0.026)
Kildare	0.0517***	0.0097
	(0.017)	(0.022)
Kilkenny	0.0443	-0.0322
	(0.031)	(0.038)
Laois	-0.0033	-0.0939***
	(0.026)	(0.031)
Leitrim	-0.0260	0.0112
	(0.039)	(0.052)
Limerick	0.0704***	-0.0077
	(0.017)	(0.021)
Longford	0.0317	0.0955**
	(0.031)	(0.039)
Louth	0.0181	-0.0078
	(0.019)	(0.023)
Mayo	0.0243	-0.0774***
	(0.021)	(0.026)

Table 4.1: Exits to the Labour Market at 12 Months (Marginal Effects) (continued)

	Male	Female
Location (Ref: Dublin)		
Meath	0.0509**	-0.0093
	(0.024)	(0.030)
Monaghan	-0.0067	-0.0384
	(0.029)	(0.035)
Offaly	0.0271	-0.0687**
	(0.025)	(0.031)
Roscommon	0.0269	-0.0638
	(0.036)	(0.047)
Sligo	-0.0526	0.0035
	(0.032)	(0.039)
Tipperary	0.0657***	0.0262
	(0.018)	(0.023)
Waterford	0.0171	-0.0105
	(0.020)	(0.026)
Westmeath	-0.0449*	-0.0742***
	(0.024)	(0.029)
Wexford	0.0437**	-0.0566**
	(0.020)	(0.024)
Wicklow	0.0377*	0.0185
	(0.021)	(0.026)
Pseudo R2	0.100	0.109
Observations	23,555	14,940

The predictive power of the models, as measured by the pseudo R^2, was satisfactory and compares well with the statistics from the earlier profiling study (see O'Connell *et al.*, 2009), which was based on a much more comprehensive dataset, that ranged from 0.11 to 0.13. If we take individuals with a predicted probability above 0.5 as likely to exit to the labour market before 52 weeks (i.e. a leaver) and those with a predicted probability below or equal to 0.5 as likely to remain on the Live Register (i.e. a stayer), overall both the male and female models will correctly identify 66 per cent of cases. The proportion of correctly predicted observations is again only marginally below that of the previous PEX study, which correctly

predicted 69 per cent of cases, and outperforms existing country profiling systems on which information is available (see O'Connell *et al.*, 2009).

Finally, Figures 4.1 and 4.2 show the distribution of predicted labour market closure probabilities among both males (Figure 4.1) and females (Figure 4.2). The male distribution is bimodal in nature, whereas the female distribution is more normal.

Figure 4.1: Probability of a Labour Market Closure: Males

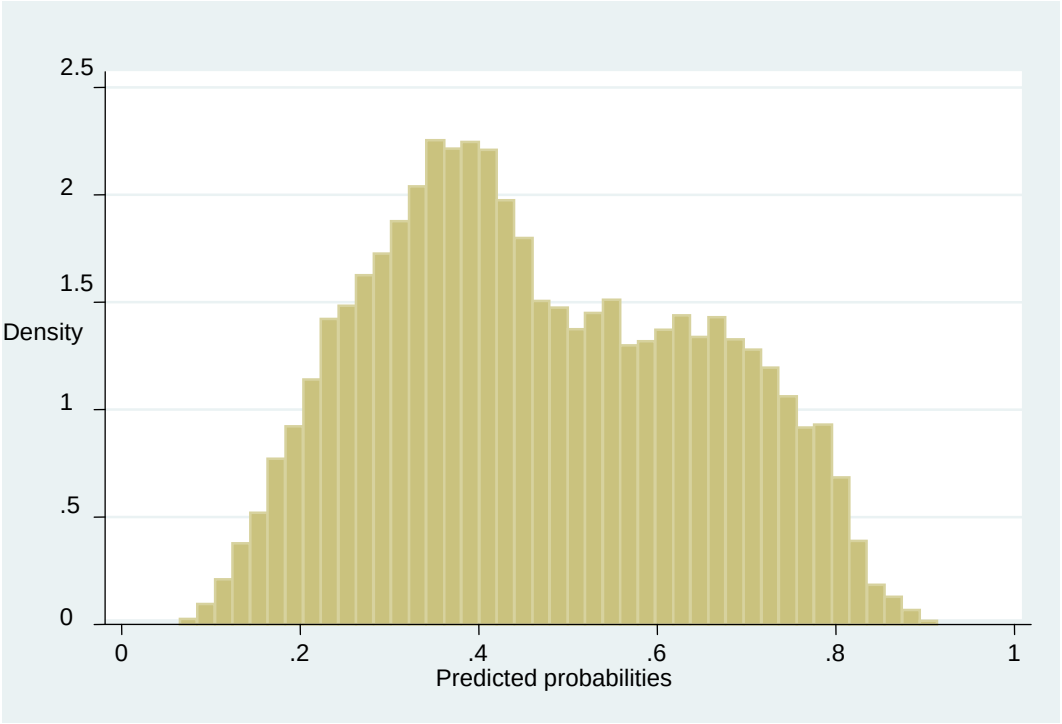
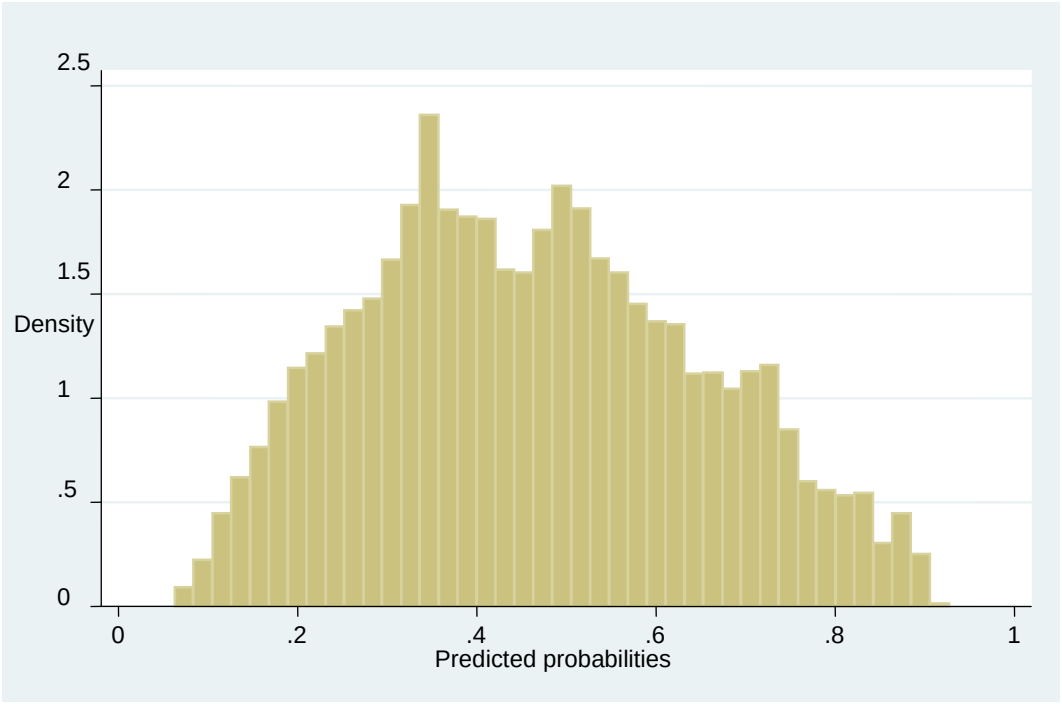


Figure 4.2: Probability of a Labour Market Closure: Females



5. Summary and Conclusions

This report summarises the research aimed at developing a profiling model, based on administrative data, that can be applied to the existing population of un-profiled claimants. As the model will be applied to a cohort of individuals who, for the most part, will have already been on the Live Register for 12 months, the tool should be thought of as a measure of Labour Market Disadvantage (LMD) as opposed to a PEX score. The LMD will provide a measure of the ex-ante probability that a person should have exited the Register before 12 months given their observable characteristics. While it is arguable that the LMD model is in some way inferior to the PEX version on the basis that it is estimated on fewer characteristics, the LMD score has the advantage of being based on more recent data relative to the PEX and will therefore more heavily reflect current labour market conditions. Furthermore, the model diagnostics indicate the the LMD models perform almost as well, in terms of predictive power. This suggests that the additional variables in the LMD model relating to previous occupation and presence in the CCS data are proxying many of the labour market impacts measured directly in the PEX study by variables such as educational attainment, unemployment and unemployment history. In conclusion, the current LMD model performs extremely well given the data constraints, and it will enable the Department to partition the un-profiled claimant population in a manner that far exceeds what would be achieved through a random draw.

References

O'Connell, P.J., S. McGuinness, E. Kelly and J Walsh. (2009). National Profiling of the Unemployed in Ireland. ESRI Research Series Number 10. Dublin: Economic and Social Reaearch Institute.

The Economic & Social Research Institute
Whitaker Square
Sir John Rogerson's Quay
Dublin 2, Ireland
+ 353 1 863 2000 www.esri.ie
ISBN 978 0 7070 03696